**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**Факультет прикладной математики-процессов управления**

**Программа бакалавриата**

**“Большие данные и распределенная цифровая платформа”**

**ОТЧЕТ**

**по лабораторной работе №2**

**по дисциплине «Функциональное программирование»**

**на тему**

**«****Анализ космических данных с использованием параллельных вычислений»**

**Вариант 1**

**Студент гр. 23Б16-пу**

**Аникин Р. С.**

**Преподаватель**

**Киямов Ж. У.**

**Санкт-Петербург**

**2024 г.**

**Оглавление**

[**Цель работы** 2](#_Toc185324406)

[**Описание задачи (формализация задачи)** 2](#_Toc185324407)

[**Теоретическая часть** 4](#_Toc185324408)

[**Описание программы** 5](#_Toc185324409)

[**Рекомендации пользователю** 7](#_Toc185324410)

[**Примечания:** 8](#_Toc185324411)

[**Рекомендации программисту** 8](#_Toc185324412)

[**Исходный код программы** 8](#_Toc185324413)

[**Контрольный пример** 8](#_Toc185324414)

[**Источники** 10](#_Toc185324415)

# **Цель работы**

Цель данной работы заключается в разработке эффективной программы для анализа больших объемов данных, получаемых с космического телескопа, с использованием методов параллельных вычислений. Существующие методы анализа изображений могут быть недостаточно быстрыми для обработки сотен тысяч космических изображений, содержащих миллионы пикселей. Поэтому в рамках данного проекта необходимо создать систему, способную одновременно обрабатывать множество изображений, обеспечивая тем самым ускорение анализа и сбор статистики о выявленных астрофизических объектах.

# **Описание задачи (формализация задачи)**

Анализ космических изображений включает следующие основные этапы:

1. **Входные данные:**
   * Набор изображений космических объектов в форматах PNG или JPG, представленных как матрицы пикселей.
2. **Обработка изображений:**  
   Для каждого изображения выполняются следующие шаги:  
   **а)** Преобразование в оттенки серого — упрощает дальнейший анализ за счёт уменьшения информации.  
   **б)** Размытие — фильтр Гаусса устраняет шумы, повышая качество анализа.  
   **в)** Бинаризация — преобразует изображение в чёрно-белый формат для выделения объектов.  
   **г)** Выделение контуров — позволяет обнаружить границы объектов на изображении.
3. **Анализ объектов:**  
   Выделенные объекты проходят анализ для определения ключевых параметров:  
   **а)** Площадь — вычисляется через функцию cv2.contourArea.  
   **б)** Вычисление координат центра объекта.  
   **в)** Яркость — сумма интенсивностей пикселей внутри контура.  
   **г)** Классификация — на основе площади и яркости объекту присваивается тип: звезда, планета, большая звезда или галактика.
4. **Параллельная обработка:**  
   Для повышения производительности каждая часть изображения обрабатывается в отдельных потоках или процессах.  
   **а)** Изображение делится на сегменты, которые обрабатываются параллельно.  
   **б)** Используются механизмы синхронизации, предотвращающие конфликты данных.
5. **Сбор статистики:**  
   Для каждого объекта собираются следующие данные:
   * Имя изображения.
   * Номер сегмента изображения.
   * Координаты центра.
   * Яркость.
   * Площадь.
   * Тип объекта.
6. **Вывод результатов:**
   * Результаты сохраняются в Excel-файл.
   * Создаются визуализации с отмеченными обнаруженными объектами.

# **Теоретическая часть**

#### **Обработка изображений:**

* **Преобразование в оттенки серого:**  
  Уменьшение избыточную информацию, сохраняя основные данные для анализа.
* **Размытие (Гауссов фильтр):**  
  Сглаживание шумы, повышая качество выделения объектов. Особенно полезно для изображений с дефектами или помехами.
* **Бинаризация:**  
  Преобразование изображение в двуцветный формат, выделяя объекты на фоне. Используются пороговые значения, например, 180, (из 255 максимально).
* **Выделение контуров:**  
  Выделение границ объектов с помощью алгоритма cv2.findContours.

#### **Классификация объектов:**

Классификация космических объектов основана на анализе физических характеристик:

* **Яркость:**  
  Используется для определения типа объекта.  
  Пример: яркие объекты с высокой интенсивностью света — звёзды или галактики.
* **Площадь:**  
  Позволяет разделить малые и крупные объекты, например, звёзды и большие звёзды.
* **Типы объектов:**
  + **Звезда:** маленький объект с высокой яркостью.
  + **Планета:** объект с большой площадью, но низкой яркостью.
  + **Большая звезда:** особо яркий объект.
  + **Галактика:** очень крупный и яркий объект.

#### **Параллельные вычисления:**

Для ускорения анализа используется многопоточная или многопроцессорная обработка:

* **Параллелизм задач:**  
  Каждая часть изображения обрабатывается независимо.
* **Используемые библиотеки:**
  + threading — для многопоточности.
  + multiprocessing — для независимой обработки данных.
* **Преимущества:**
  + Ускорение обработки.
  + Возможность анализа больших изображений.
* **Проблемы:**
  + **Гонки данных:** решаются с помощью блокировок и очередей.
  + **Перегрузка процессора:** оптимизация количества потоков для числа доступных ядер.

**Основные шаги программы**

1) Загрузка и предварительная обработка изображения.

2) Обнаружение контуров объектов в параллельном режиме.

3) Анализ и классификация объектов в параллельном режиме.

4) Сохранение результатов анализа.

5) Визуализация результатов.

6) Завершение программы

# **Описание программы**

Программная реализация написана на языке Python 3.10 с использованием следующих библиотек: cv2 для работы с изображениями, multiprocessing для многопоточной обработки, numpy для математических операций, openpyxl для работы с Excel-файлами, и tkinter для создания графического интерфейса. Программа организована в модуле, который фокусируется на обработке астрономических изображений, выделении контуров объектов и сохранении данных в Excel-файлы.

В процессе разработки программы использовалось 15 функций, каждая из которых имеет четко определенное назначение:

Таблица 1. parallel\_processing\_space.py

|  |  |  |
| --- | --- | --- |
| **Функция** | **Описание** | **Возвращаемое значение** |
| create\_interface() | Создание графического интерфейса с кнопками для выбора изображений, пути сохранения и начала анализа. | Не возвращает |
| Choosephotos() | Открытие диалогового окна для выбора директории с изображениями для анализа. | Не возвращает |
| savePath() | Открытие диалогового окна для выбора директории сохранения Excel-файла с результатами анализа. | Не возвращает |
| Analyze() | Основная функция анализа изображений. Вызывает обработку всех изображений, сохранение результатов и уведомление пользователя о завершении процесса. | Не возвращает |
| processAllphotos() | Обработка всех изображений из указанной директории. | Не возвращает |
| Processphoto() | Разбиение изображения на части, анализ каждой части изображения, объединение данных о всех обнаруженных объектах. | allObjectsData – список объектов |
| Splitphoto() | Разделение исходного изображения на части для параллельной обработки. | photoParts список |
| analyzePart() | Анализ отдельной части изображения: выделение контуров, вычисление площади, яркости объектов и их классификация. | objectsData (список словарей), contourCenters (список) |
| сlassification() | Классификация объектов по площади и яркости на основе заданных критериев (звезда, яркая звезда, планета). | Object\_type: string |
| savephotoPart() | Сохранение частей изображений с выделенными контурами объектов в отдельные файлы. | Не возвращает |
| Save() | Сохранение всех результатов анализа в Excel-файл с возможностью настройки ширины столбцов. | Не возвращает |

## **Рекомендации пользователю**

 **Запуск программы**

* Дважды щёлкните по исполняемому файлу, чтобы открыть графический интерфейс программы.

 **Выбор изображений для анализа**

* Нажмите кнопку **"Выбрать изображения"**.
* В появившемся окне выберите папку, где находятся изображения для обработки. Поддерживаются форматы .jpg и .png.
* После выбора папки путь к ней отобразится в строке **"Выбраны изображения"**.

 **Указание папки для сохранения результатов**

* Нажмите кнопку **"Сохранить путь"**.
* В открывшемся окне выберите папку для сохранения итогового файла с результатами (формат Excel).
* Путь сохранения будет отображён в строке **"Сохранить в"**.

 **Запуск анализа изображений**

* Нажмите кнопку **"Анализ"**. Программа начнёт обработку изображений, выполняя анализ каждого фрагмента и выделяя объекты (например, звёзды или планеты).
* Продолжительность обработки зависит от количества и размера изображений.

 **Завершение анализа**

* После завершения обработки появится уведомление, подтверждающее успешное сохранение результатов в выбранный Excel-файл.

 **Просмотр результатов**

* Откройте Excel-файл (например, statistic.xlsx) в выбранной папке. Таблица содержит следующую информацию:
  + Название изображения.
  + Номер сегмента.
  + Координаты объектов.
  + Яркость объекта.
  + Площадь объекта.
  + Тип объекта (звезда, большая звезда, планета, галактика).

 **Дополнительно**

* Обработанные изображения сохраняются в папке **photo\_parts**. Каждое изображение разбито на фрагменты, где выделены найденные объекты.

## **Рекомендации программисту**

* Убедитесь, что установлены все необходимые (OpenCV, NumPy, OpenPyXL, Tkinter) библиотеки.
* Убедитесь, что все изображения находятся в одной папке (автоматически сохраняются именно туда)

# **Исходный код программы**

[**https://github.com/RomanMuteki/func\_prog**](https://github.com/RomanMuteki/func_prog)

# **Контрольный пример**

1. Запуск программы: Для запуска программы используйте файл **main.py.** Программа запустит графический интерфейс (Рис. 1), в котором можно будет выбрать “Выбрать изображения” или “Сохранить путь” и “Анализ”.

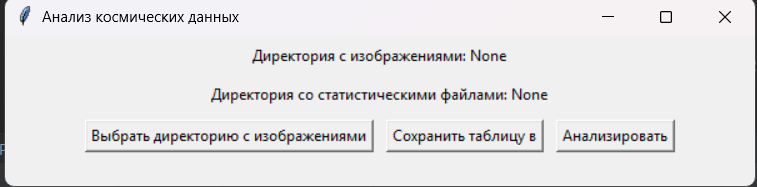


Рис 1. Графический интерфейс

2. Выбрать изображения: При нажатии на кнопку: “Выбрать директорию с изображениями” откроется проводник в котором будет необходимо выбрать папку из которой мы будем выбирать изображения для анализа.

3. Путь сохранения: При нажатии на кнопку: “ Сохранить таблицу в” откроется проводник в котором будет необходимо выбрать папку в которую мы впоследствии будем сохранять получившиеся результаты.

4. Анализ: При нажатии на кнопку: “Анализировать” запустится процесс обработки картинок из выбранной папки, создастся папка: photo\_parts в которой будут части изображения и обведённые на них объекты, а также создастся .xlsx файл в котором будет также представлена информация об этих объектах.

5. Итог: После завершения анализа появится окно(Рис. 2), уведомляющее о том, куда сохранены результаты.

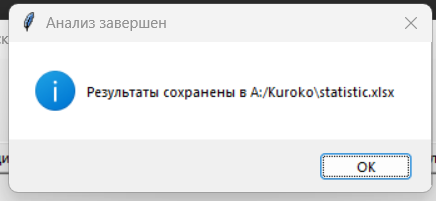


Рис 2. Окно завершения

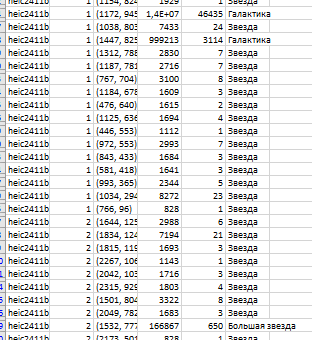


Рис 3. Файл statistic.xlsx

# **Источники**

* **OpenCV**

<https://docs.opencv.org/4.x/>

*дата обращения: (19.11.2024)*

* **NumPy**

[*https://numpy.org/*](https://numpy.org/)

*дата обращения: (19.11.2024)*

* **OpenPyXL**

[https://openpyxl.readthedocs.io/en/stable/](https://openpyxl.readthedocs.io/en/stable/%20)

*дата обращения: (19.11.2024)*